

AMENDMENT TO THE SPECIFICATION

Please amend the paragraph beginning on page 10, line 25 as follows:

Compression module 302 can be embodied in any system capable of executing one or~~ore~~ more of the query log compression operations set forth below. Accordingly, compression module 302 can be computer hardware, such as that set forth above, computer software embodied in any suitable programming language, or any combination of the two. FIG. 2 also illustrates dashed line 306 in order to illustrate that the compressed query log 304 can be fed back to compression module 302 in order to take multiple compression passes on the log.

Please amend the paragraph beginning on page 16, line 12 as follows:

Regardless of the manner in which the blocking filters are applied, the process of identifying an "extra" term when comparing each, for example, length five query with all length four queries should be done efficiently. A simple comparison between two strings, whether at character level (two words compared letter-by-letter as in spelling correction) or at token level (two phrases compared word-by-word) is called "edit distance." When the edit distance of token comparison of two query strings is exactly one, that means that only one token difference has been identified and accordingly that subsumption is possible. Preferably, the query strings are sorted by word tokens prior to the subsumption process, as part of an ordinary normalization process and such sorting is preferably a precondition for edit distance computation. Since edit distance computation can be very computationally intensive when applied to logs potentially containing millions of queries, subsumption preferably employ~~employees~~ a shortcut. Specifically, for subsumption, full edit-distance computation is only attempted when a one-difference could in principle occur for the two strings being compared. Accordingly, a recheck for possible one-difference comparisons is generated using an index over all of the "short" strings by their first and second token words.